

PATENT

VERTICAL MOS TRANSISTOR

BACKGROUND OF THE INVENTION

5 1. Field of the Invention.

The present invention relates to MOS transistors and, more particularly, to a vertical MOS transistor.

10 2. Description of the Related Art.

A MOS transistor is a well-known semiconductor device that controls the magnitude of a current that flows through the transistor. The current typically flows horizontally, such as from the drain to the
15 source of an NMOS transistor. MOS transistors can be used as both low-power and high-power devices.

BRIEF DESCRIPTION OF THE DRAWINGS

20 FIG. 1A is a plan view illustrating an example of a vertical MOS transistor 100 in accordance with the present invention. FIG. 1B is a cross-sectional view of transistor 100 taken along line 1B-1B of FIG. 1A.

FIG. 2 is a cross-sectional view illustrating an example of a vertical MOS transistor 200 in accordance with the present invention.

25 FIG 3A is a plan view illustrating an example of a vertical MOS transistor array 300 in accordance with the present invention. FIG. 3B is a cross-sectional view of array 300 taken along line 3B-3B of FIG. 3A.

FIG. 4 is a cross-sectional view illustrating an example of a vertical MOS transistor array 400 in accordance with the present
30 invention.

FIGs. 5A-5G are a series of cross-sectional views illustrating an example of a method of forming a vertical MOS transistor in accordance with the present invention.

FIGs. 6A-6D are a series of cross-sectional views illustrating an example of a method of forming a vertical MOS transistor in accordance with the present invention.

DETAILED DESCRIPTION OF THE INVENTION

FIG. 1A shows a plan view that illustrates an example of a vertical MOS transistor 100 in accordance with the present invention. FIG. 1B shows a cross-sectional view of transistor 100 taken along line 1B-1B of FIG. 1A. As described in greater detail below, transistor 100 provides a vertical current flow.

As shown in FIG. 1B, vertical MOS transistor 100 includes a p-silicon region 110, and an n+ region 112 that contacts the top surface of p-silicon region 110. P-silicon region 110 is formed to have a crystallographic orientation of $\langle 110 \rangle$. In addition, p-silicon region 110 can be formed to have a depth of, for example, several hundred micrometers, while n+ region 112 can be formed to have a depth of, for example, one micrometer ($1\mu\text{M}$).

As shown in FIGs. 1A-1B, transistor 100 includes an epitaxial layer 114 that is formed on n+ region 112. Epitaxial layer 114, which also has a crystallographic orientation of $\langle 110 \rangle$, includes a p-region 120 that contacts n+ region 112, and an overlying n+ region 122 that contacts p-region 120. Epitaxial layer 114 can be formed to have a depth of, for example, three micrometers ($3\mu\text{M}$). For example, p-region 120 can have a depth of two micrometers ($2\mu\text{M}$), while n+ region 122 can have a depth of one micrometer ($1\mu\text{M}$).

In addition to the above, transistor 100 also includes a trench 130 that extends from the top surface of n+ region 122 through n+ region 122, p- region 120, and partially into n+ region 112. Further, trench 130 is lined with a layer of insulation material 132, such as gate oxide, and filled with a conductive gate region 134. Conductive gate region 134 can be implemented with, for example, an n-type polysilicon region that has been predoped to achieve an appropriate work function difference to silicon, thereby achieving a practical threshold voltage.

As further shown in FIGs. 1A-1B, a layer of isolation material 140, such as a layer of oxide, can be formed over n+ region 122, insulation material layer 132 and gate region 134. In addition, a metallic contact 142 can be formed through isolation material 140 to make an electrical connection with gate region 134, and a metallic contact 144 can be formed through isolation material 140 to make an electrical connection with n+ region 122 on a first side of gate region 134.

Further, as shown by the dashed lines in FIGs. 1A-1B, a metallic contact 146 can optionally be formed through isolation material 140 to make an electrical connection with n+ region 122 on a second side of gate region 134 that lies opposite the first side. Additional metallic contacts can also be formed through isolation material 140 to make electrical connections with n+ region 122 and gate region 134.

Although not shown, a conventionally-formed interconnect structure, which includes metal traces and vias, is formed on contacts 142, 144, and 146 (and any additional contacts) such that a gate voltage can be applied to gate region 134 via contact 142 (and any additional contacts to gate region 134), and a source voltage can be applied to n+ region 122 via contact 144 (and any additional contacts to n+ region 122).

As additionally shown in FIGs. 1A-1B, transistor 100 includes a metallic plug 150, such as a copper or other metallic plug, that is formed

through p- silicon region 110 to make an electrical connection with n+ region 112. Although not shown, metallic plug 150 is conventionally connected so that a drain voltage can be applied to metallic plug 150.

5 In operation, transistor 100 operates like a conventional MOS transistor except that current flows vertically adjacent to the sides of gate region 134. When a voltage difference exists between n+ region 122 and n+ region 112 and the gate voltage (which is applied to gate region 134) is equal to or greater than the threshold voltage, the channel region of p- region 120 that lies adjacent to gate region 134
10 inverts and a current flows from plug 150 to n+ region 112 to p- region 120 to n+ region 122 and on to contact 144.. When the gate voltage falls below the threshold voltage, the current stops (except for a leakage current).

FIG. 2 shows a cross-sectional view that illustrates an example of
15 a vertical MOS transistor 200 in accordance with the present invention. FIG. 2 is taken along line 1B-1B of FIG. 1A. Transistor 200 is similar to transistor 100 and, as a result, utilizes the same reference numerals to designate the structures which are common to both transistors.

As shown in FIG. 2, transistor 200 differs from transistor 100 in
20 that transistor 200 includes an insulator layer 210, such as a layer of oxide, that is formed between p- silicon region 110 and n+ region 112. Insulator layer 210 can represent the insulator layer in a silicon-on-insulator (SOI) structure.

One of the advantages of transistors 100 and 200 is that
25 transistors 100 and 200 have a compact footprint which, in turn, allows a large number of transistors to be formed in parallel to form a high current, high power MOS transistor array in a relatively small surface area.

FIG 3A shows a plan view that illustrates an example of a vertical
30 MOS transistor array 300 in accordance with the present invention. FIG.

3B shows a cross-sectional view of array 300 taken along line 3B-3B of FIG. 3A. As shown in FIG. 3B, transistor array 300 includes a p- silicon region 310, and an n+ region 312 that contacts the top surface of p- silicon region 310.

5 In addition, array 300 includes an epitaxial layer 314 that is formed on n+ region 312. Epitaxial layer 314, includes a p- region 320 that contacts n+ region 312, and an overlying n+ region 322 that contacts p- region 320. As above, p- silicon region 310 and epitaxial layer 314 are formed to have a crystallographic orientation of $\langle 110 \rangle$.
10 Further, regions 310, 312, 314, 320, and 322 can be the same as regions 110, 112, 214, 120, and 122 shown in FIGs. 1A and 1B.

 Array 300 also includes a series of trenches TR1-TRn that each extend from the top surface of n+ region 322 through n+ region 322, p- region 320, and partially into n+ region 312. Further, array 300 includes
15 a corresponding series of isolation regions IS1-ISn that line the trenches TR1-TRn, and a corresponding series of n+ conductive gate regions PY1-PYn that contact the isolation regions IS1-ISn, respectively, to fill the trenches TR1-TRn, respectively. As above, the conductive gate regions PY1-PYn can be implemented with, for example, n-type
20 polysilicon that has been predoped to achieve an appropriate work function difference to silicon, thereby achieving a practical threshold voltage.

 As also shown in FIGs. 3A-3B, a layer of insulation material 340, such as a layer of oxide, can be formed over n+ region 322, isolation
25 regions IS1-ISn, and gate regions PY1-PYn. In addition, a series of gate contacts GC1-GCn can be formed through insulation material 340 to make an electrical connection with the gate regions PY1-PYn, and a series of source contacts SC1-SCm can be formed through insulation material 340 to make electrical connections with n+ region 322 adjacent
30 to the gate regions PY1-PYn.

Additional metallic contacts can also be formed through insulation material 340 to make electrical connections with n+ region 322 and the gate regions PY1-PYn. Further, the arrangement of the gate and source contacts GC1-GCn and SC1-SCm are shown only for purposes of
5 illustration. Other arrangements can also be used.

Although not shown, a conventionally-formed interconnect structure, which includes metal traces and vias, is formed on the gate and source contacts GC1-GCn and SC1-SCm such that a gate voltage can be applied to the gate contacts GC1-GCn, and a source voltage,
10 such as ground, can be applied to the source contacts SC1-SCm.

As additionally shown in FIGs. 3A-3B, vertical MOS transistor array 300 includes a metallic plug 350, such as a copper or other metallic plug, that is formed through p- silicon region 310 to make an electrical connection with n+ region 312. Although not shown, a
15 conventionally-formed interconnect structure is formed on metallic plug 350 such that a drain voltage can be applied to metallic plug 350.

In operation, array 300 operates like MOS transistor 100 except that current flows vertically adjacent to the sides of each gate region PY1-PYn. When a voltage difference exists between n+ region 322 and
20 n+ region 312 and the gate voltage (which is applied to gate regions PY1-PYn) is equal to or greater than the threshold voltage, the channel regions of p- region 320 that lie adjacent to each gate region PY1-PYn invert and current flows from plug 350 to n+ region 312 to p- region 320 to n+ region 322 and on to source contacts SC1-SCm. When the
25 gate voltage falls below the threshold voltage, the current stops (except for a leakage current).

FIG. 4 shows a cross-sectional view that illustrates an example of a vertical MOS transistor array 400 in accordance with the present invention. FIG. 4 is taken along line 3B-3B of FIG. 3A. Array 400 is
30 similar to array 300 and, as a result, utilizes the same reference

numerals to designate the structures which are common to both transistors.

As shown in FIG. 4, array 400 differs from array 200 in that array 400 includes an insulator layer 410, such as a layer of oxide, that is
5 formed between p- silicon region 310 and n+ region 312. Insulator layer 310 can represent the insulator layer in a silicon-on-insulator (SOI) structure.

One of the advantages of arrays 300 and 400 is that arrays 300 and 400 have low drain-to-source resistance turn on values (R_{dson}) due
10 to plug 350 which provides the majority of the current path. In addition, plug 350 in arrays 300 and 400 reduces the amount of die surface area that is required by the transistors in a power array application.

Another advantage of arrays 300 and 400 of the present invention is that by creating trenches in the form of an array, the
15 vertical MOS channels are controlled by a surrounding electric field. As a result, the intrinsic performance of the transistors in the present invention provides an increased sub-threshold performance capability. Thus, leakage versus drive current trade off performance can be better than a conventional MOS power transistor.

20 FIGs. 5A-5G show a series of cross-sectional views that illustrate an example of a method of forming a vertical MOS transistor in accordance with the present invention. As shown in FIG. 5A, the method, which utilizes a p- single crystal silicon region 510 that has a $\langle 110 \rangle$ crystallographic orientation, includes the step of implanting the
25 surface of silicon region 510 with an n-type dopant to form an n+ region 512. (Peripheral areas of region 510 which are not to be implanted are masked.)

Once n+ region 512 has been formed, a p- silicon region 514 is epitaxially grown on region 512 to have a depth of, for example, two
30 micrometers ($2\mu\text{M}$), and an n+ silicon region 516 is epitaxially grown on

region 514 to have a depth of, for example, one micrometers ($1\mu\text{M}$). Alternately, to minimize stacking faults, a p- region three micrometers ($3\mu\text{M}$) can be epitaxially grown, followed by an n+ implant that forms n+ region 516 and defines p- region 514. Regions 514 and 516 both
5 have a $\langle 110 \rangle$ crystallographic orientation. Next, a hard mask 518, such as a nitride mask, is formed and patterned on the top surface of n+ silicon region 516.

After this, as shown in FIG. 5B, the exposed regions of n+ silicon region 516 and the underlying regions are etched with a wet etchant,
10 such as KOH (ethanol) or TMAH (tetramethylammonium hydroxide), to form a series of trenches TR1-TRn that each expose n+ region 512. A KOH etch is a relatively slow etch (e.g., 60 microns/hour), but when used to etch a $\langle 110 \rangle$ region of silicon, produces a series of very deep trenches, e.g., greater than three micrometers ($>> 3\mu\text{M}$) with
15 substantially anisotropic side walls. Due to the time required to complete an etch, a production process would etch a number of wafers at the same time. Following the etch, mask 518 is removed.

Next, as shown in FIG. 5C, a layer of insulation material 520, such as oxide, is formed on n+ region 512, p- region 514, and n+
20 region 516 to line the trenches TR1-TRn. Following this, a layer of conductive material 522 is formed on insulation layer 520 to fill up the trenches TR1-TRn. Conductive material 522 can be implemented with, for example, an n-type polysilicon material that has been predoped to achieve an appropriate work function difference to silicon, thereby
25 achieving a practical threshold voltage.

As shown in FIG. 5D, once conductive layer 522 has been formed, conductive layer 522 and insulation layer 520 are planarized until insulation layer 520 has been removed from the surface of n+ region 516. The planarization step forms a series of insulation regions
30 IN1-INn that line the trenches TR1-TRn, respectively, and a series of

conductive gate regions PY1-PYn that fill the trenches TR1-TRn, respectively.

After this, a layer of insulator material 524 is formed on n+ region 516, insulation regions IN1-INn, and conductive gate regions PY1-PYn. Next, a mask 526 is formed and patterned on insulator layer 524. The exposed regions of insulator layer 524 are then etched to expose the surfaces of n+ region 516 and gate regions PY1-PYn. The etch forms a number of contact openings 530. Mask 526 is then removed.

Following this, a layer of conductive material, which can include a diffusion barrier, is formed on insulator layer 524 to fill up the contact openings 530. As shown in FIG. 5E, the layer of conductive material is then planarized to remove the layer of conductive material from the top surface of insulator layer 524 and form a number of conductive contacts 532. An interconnect structure, which includes metal traces and vias, is then formed on insulator layer 524 to make electrical connections with the contacts 532.

Next, as shown in FIG. 5F, a hard mask 540, such as nitride, is formed and patterned on the bottom surface of p- silicon region 510 (which can be ground down prior to the formation of mask 540). After this, the exposed region of p- silicon region 510 is etched with a wet etchant, such as KOH or TMAH, to form a trench 542 that exposes n+ region 512.

As noted above, a KOH etch is a relatively slow etch that can produce a very deep trench of, for example, greater than twenty five micrometers ($>25\mu\text{M}$) with substantially anisotropic side walls when used to etch a $\langle 110 \rangle$ region of silicon. As further noted above, due to the time required to complete an etch, a production process would etch a number of wafers at the same time. Following the etch, mask 540 is removed and the surface is cleaned with an HF dip.

Next, as shown in FIG. 5G, a barrier layer 544 is formed on p-silicon region 510 and n+ region 512 to line trench 542. Once barrier layer 544 has been formed, a copper seed layer 546 is formed on barrier layer 544. Next, a conductive metal layer 548 is electroplated over seed layer 546 to fill up trench 542.

Conductive metal layer 548, seed layer 546, and barrier layer 544 are then planarized to remove barrier layer 544 from the bottom surface of p-silicon region 510 to form a conductive plug 550 that extends through p-silicon region 510 to make an electrical connection with n+ region 512.

FIGs. 6A-6D show a series of cross-sectional views that illustrate an example of a method of forming a vertical MOS transistor in accordance with the present invention. As shown in FIG. 6A, the method, which utilizes a p- single crystal silicon region 608 that has a <110> crystallographic orientation and an overlying layer of insulator material 610, includes the step of, in sequence, epitaxially growing an n+ region 612, a p- region 614, and an n+ region 616. Regions 612, 614 and 616 are grown to have depths of, for example, one micrometer (1μM), two micrometers (2μM), and one micrometer (1μM), respectively.

Alternately, to minimize stacking faults, regions 612, 614, and 616 can be epitaxially grown in one or more steps using a lower dopant concentration, and then implanted to form n+ regions 612 and 616. For example, a p- epitaxial layer can be grown, followed by a first implant that forms n+ region 612 and a second implant that forms n+ region 616. Alternately, an n- epitaxial layer can be grown, and then implanted to form n+ region 612. Following this, a p- epitaxial layer can be grown, and then implanted to form n+ region 616 where p- region 614 lies between regions 612 and 616.

Following this, the method continues with the steps shown in FIGs. 5B-5F, beginning with the step of forming hard mask 518 on n+

region 616, and ending with the step of forming a hard mask 540 on the bottom surface of p- silicon region 608. After this, as shown in FIG. 6B, the exposed region of p- silicon region 608 is etched with a wet etchant, such as KOH or TMAH, to form a trench 620 that exposes insulator material 610. Thus, insulator material 610 can be utilized as an etch stop.

Next, as shown in FIG. 6C, the exposed region of insulator material 610 is etched, such as with an HF dip, to expose n+ region 612. Hard mask 540 is then removed. After this, as shown in FIG. 6D, a barrier layer 622 is formed on p- silicon region 608 and n+ region 612 to line trench 620. Once barrier layer 622 has been formed, a copper seed layer 624 is formed on barrier layer 622.

Next, a conductive metal layer 626 is electroplated over the seed layer to fill up trench 620. Conductive metal layer 626, seed layer 624, and barrier layer 622 are then planarized to remove barrier layer 622 from the bottom surface of p- silicon region 608 to form a conductive plug 628 that extends through p- silicon region 608 to make an electrical connection with n+ region 612.

One of the advantages of the methods illustrated by FIGs. 5A-5G and 6A-6D is that both methods can be easily integrated into a conventional buried-layer CMOS or BiCMOS fabrication process, with the formation of copper plug 350 being the only required additional steps to the methods.

It should be understood that the above descriptions are examples of the present invention, and that various alternatives of the invention described herein may be employed in practicing the invention. For example, metallic plugs 150 and 350 can be isolated from p- silicon regions 110 and 310, respectively so that metallic plugs 150 and 350 can have a voltage that is different from the voltage on p- silicon regions 110 and 310, respectively. Thus, it is intended that the following claims

PATENT

define the scope of the invention and that structures and methods within the scope of these claims and their equivalents be covered thereby.